#Problem: Word Completion

Given the set of words in the dictionary [composed from the list of **N** words.], we want to suggest/predict the most likely word after providing some characters inputs from the keyboard.

The problem assumes there are two dictionaris:

1. Small Dictionary: contain about 1000 words.
   URL: https://github.com/first20hours/google-10000-english (Sample 1000 words using uniform distribution.)
2. Larger Dictionary:
   URL: https://www.keithv.com/software/wlist/ (wlist_match1.zip is for out application.)

#Solutions

There are many algorithms available to the auto suggestions for a word like trie based lookup (as we have implemented in *wisemd* system). I propose a solution that is based on the vector based similarity. We can use cosine similarity metric to rank the nearest word for the given incomplete one.

Tasks:

1. Write a dictionary vectorizer based on character ngrams and write a module to interact(read/write) with it.
2. Write a vectorizer for an input word with same strategy of dictionary vectorizer.
3. Measure the cosine similarity(angle) of the input vector against the dictionary entries.
4. Rank the first **n** similar words based on the score.

In our assumptions, we have two dictionaries, basically are of two different sizes. Analyze the performance of your solution to these dicionaries and list out the problems of your solutions.

#Outcome:

There should an input of partially complete word, could in text box or in console and the suggestions are listed in the console or in the text area.

You can choose any language of your choice but I prefer to write in **Clojure** or **Scala** because we are learning to write program in functional way.

Prepared by: Santa, WiseYak, 2020-07-06